WHAT IS CLAIMED IS:

1. A method of identifying one or more positions in a polymer family, the method comprising:
   (a) accessing data representing a multiple sequence alignment (MSA) of a plurality of polymer sequences; and
   (b) identifying one or more positions within the MSA that have statistically significant conservation energy values using the following equation:

$$\Delta G_i^{stat} = kT^* \sqrt{\sum_x \left(\ln \frac{P_i^x}{P_{MSA}^x} \right)^2}$$

   wherein:

   i is a position in the MSA;

   ΔG_i^{stat} is the conservation energy value for position i;

   P_i^x is the probability of monomer x at position i;

   P_{MSA}^x is the probability of monomer x in the MSA; and

   kT* is an energy unit, where k is Boltzmann's constant.

2. The method of claim 1, wherein the method is executed using a machine.

3. A program storage device readable by the machine of claim 2 and encoding instructions executable by the machine for performing the operations recited in the claim.

4. The method of claim 1, further comprising generating a graphical image of the conservation energy values.

5. The method of claim 1, wherein the polymer sequences comprise protein sequences.

6. The method of claim 1, wherein monomer x comprises amino acid x.

Sub A17

7. The method of claim 1, wherein the data accessed comprises data from the PDZ domain family.

8. The method of claim 1, wherein the data accessed comprises data from the $p21^{ras}$ domain family.

9. The method of claim 1, wherein the data accessed comprises data from the hemoglobin domain family.

10. A method of identifying one or more positions in a polymer family, the method comprising:

(a) accessing data representing a multiple sequence alignment (MSA) of a plurality of polymer sequences;

(b) calculating a conservation energy value for each position in the MSA using the following equation:

$$\Delta G_i^{stat} = kT^* \sqrt{\sum_x \left(\ln \frac{P_i^x}{P_{MSA}^x} \right)^2}$$

wherein:

i is a position in the MSA;

ΔG_i^{stat} is the conservation energy value for position i;

P_i^x is the probability of monomer x at position i;

P_{MSA}^x is the probability of monomer x in the MSA;

kT* is an energy unit, where k is Boltzmann's constant; and

(c) identifying one or more positions within the MSA that have statistically significant conservation energy values.

11. The method of claim 10, wherein the method is executed using a machine.

12. A program storage device readable by the machine of claim 11 and encoding instructions executable by the machine for performing the operations recited in the claim.

13. The method of claim 10, further comprising generating a graphical image of the conservation energy values.

14. The method of claim 10, wherein the polymer sequences comprise protein sequences.

15. The method of claim 10, wherein monomer x comprises amino acid x.

Sub A2

16. ~~The method of claim 10, wherein the data accessed comprises data from the PDZ domain family.~~

17. The method of claim 10, wherein the data accessed comprises data from the $p21^{ras}$ domain family.

18. The method of claim 10, wherein the data accessed comprises data from the hemoglobin domain family.

19. A method useful in identifying interacting monomers in a polymer family, the method comprising:

(a) accessing data representing a multiple sequence alignment (MSA) of a plurality of polymer sequences;

(b) calculating a respective conservation energy value for each position in the MSA using the following equation:

$$\Delta G_i^{stat} = kT^* \sqrt{\sum_x \left(\ln \frac{P_i^x}{P_{MSA}^x} \right)^2}$$

wherein:

i is a position in the MSA;

ΔG_i^{stat} is the conservation energy value for position i;

P_i^x is the probability of monomer x at position i;

P_{MSA}^x is the probability of monomer x in the MSA;

kT* is an energy unit, where k is Boltzmann's constant;

(c) perturbing a position in the MSA other than position i;

(d) re-calculating the respective conservation energy value for each position in the MSA to yield a perturbed conservation energy value; and

(e) identifying positions within the MSA that have statistically significant differences between their respective conservation energy values and their perturbed conservation energy values.

20. The method of claim 19, wherein the perturbing includes:

selecting a position j in the MSA; and

selecting a subset of the MSA, the subset having one or more monomers at position j in the MSA.

21. The method of claim 20, wherein the re-calculating and identifying include:

for each position in the MSA, calculating a vector difference $\Delta\Delta G^{stat}$ between the conservation energy value of the MSA and a conservation energy value of the subset of the MSA using the following equation:

$$\Delta\Delta G_{i,j}^{stat} = kT^* \sqrt{\sum_x \left(\ln \frac{P_{i|\delta j}^x}{P_{MSA|\delta j}^x} - \ln \frac{P_i^x}{P_{MSA}^x} \right)^2}$$

wherein:

$\Delta\Delta G_{i,j}^{stat}$ is the vector difference in conservation energy values for position i;

$P_{i|\delta j}^x$ is the probability of monomer x at position i of the subset;

$P_{MSA|\delta j}^x$ is the probability of monomer x in the subset; and

identifying positions within the MSA that have statistically significant $\Delta\Delta G^{stat}$ values.

22. The method of claim 21, further comprising generating a graphical image of the $\Delta\Delta G^{stat}$ values.

23. The method of claim 19, wherein the method is executed using a machine.

24. A program storage device readable by the machine of claim 23 and encoding instructions executable by the machine for performing the operations recited in the claim.

25. The method of claim 19, wherein the polymer sequences comprise protein sequences.

26. The method of claim 19, wherein monomer x comprises amino acid x.

27. The method of claim 19, wherein the data accessed comprises data from the PDZ domain family.

28. The method of claim 19, wherein the data accessed comprises data from the $p21^{ras}$ domain family.

29. The method of claim 19, wherein the data accessed comprises data from the hemoglobin domain family.

30. A machine-executed method of quantitatively identifying interacting amino acids in a protein family, the method comprising:

(a) accessing data representing a multiple sequence alignment (MSA) of a plurality of protein sequences that are members of a common structural family;

(b) for each position in the MSA, calculating a respective conservation energy value using the following equation:

$$\Delta G_i^{stat} = kT^* \sqrt{\sum_x \left(\ln \frac{P_i^x}{P_{MSA}^x} \right)^2}$$

wherein:

i is a position in the MSA;

ΔG_i^{stat} is the conservation energy value for position i;

P_i^x is the probability of amino acid x at position i;

P_{MSA}^x is the probability of amino acid x in the MSA;

kT* is an energy unit, where k is Boltzmann's constant;

(c) selecting a position j in the MSA;

(d) selecting a subset of the MSA, wherein the subset has one or more amino acids at position j in the multiple sequence alignment;

(e) for each position in the multiple sequence alignment, calculating a vector difference between the respective conservation energy value of the multiple sequence alignment and the respective conservation energy value of the subset of the multiple sequence alignment; and

(f) identifying positions within the MSA that have statistically significant vector differences.

31. A method of analyzing data comprising:

(a) providing at least one protein having a crystal structure and multiple positions;

(b) solving the crystal structure of the at least one protein; and

(c) identifying pathways between interacting positions on the at least one protein.

32. A method of analyzing the effect of perturbation on a protein, comprising:

(a) accessing data representing at least one protein and at least one perturbed protein, both proteins having at least one identical atom;

(b) calculating a quantity of change Δ_{struct} to the atom using the following equation:

$$\Delta_{struct} = \frac{|\vec{r}_{mut}|}{\sqrt{\sigma_{mut}^2 + \sigma_{wt}^2}}$$

wherein:

$|\vec{r}_{mut}|$ is the magnitude of a vector connecting the position of the atom in the at least one perturbed protein and the position of the atom in the at least one protein;

σ_{mut} is a standard deviation of the atom in the at least one perturbed protein; and

σ_{wt} is a standard deviation of the atom in the at least one protein.

33. A method of analyzing data, comprising:

(a) accessing data representing at least one protein, a first perturbation of the at least one protein yielding a first perturbed protein, a second perturbation of the at least one protein yielding a second perturbed protein, and a double perturbation of the at least one protein yielding a double perturbed protein, the double perturbation comprising both the first and second perturbations, the proteins each having at least one identical atom;

(b) calculating a quantity of structural coupling $\Delta\Delta_{struct}$ between the first and second perturbations using the following equation:

$$\Delta\Delta_{struct} = \frac{|\vec{r}_{mut1} - \vec{r}_{mut1|mut2}|}{\sqrt{\sigma_{wt}^2 + \sigma_{mut1}^2 + \sigma_{mut2}^2 + \sigma_{mut1,mut2}^2}}$$

wherein:

$\vec{r}_{mut1}$ is a vector connecting the position of the atom in the first perturbed protein and the position of the atom in the at least one protein;

$\vec{r}_{mut1|mut2}$ is a vector connecting the position of the atom in the double perturbed protein and the position of the atom in the second perturbed protein;

σ_{wt} is a standard deviation of the atom in the at least one protein;

σ_{mut1} is a standard deviation of the atom in the first perturbed protein;

σ_{mut2} is a standard deviation of the atom in the second perturbed protein; and

$\sigma_{mut1,mut2}$ is a standard deviation of the atom in the double perturbed protein.

34. A method of analyzing microarray data comprising:

(a) accessing microarray data representing an expression level of at least one gene, an expression level of the at least one gene resulting from a first perturbation, an expression level of the at least one gene resulting from a second perturbation, and an expression level of the at least one gene resulting from a double perturbation comprising both the first and second perturbations; and

(b) calculating a degree of coupling $\Delta\Delta E$ between the first and second perturbations using the following equation:

$$\Delta\Delta E = kT^{*} \ln\left(\frac{f_1}{f_2}\right)$$

wherein:

f_1 is the fold effect of the gene due to the first perturbation relative to the at least one gene;

f_2 is the fold effect of the gene due to the double perturbation relative to the second perturbation; and

kT' is an energy unit, where k is Boltzmann's constant.